REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
02-03-2010	Final Report	1-Jul-2007 - 30-Nov-2009

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Using the Maximum Entropy Principle as a Unifying Theory for Characterization and Sampling of Multi-scaling Processes in Hydrometeorology	W911NF-07-1-0126
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
	611102

6. AUTHORS	5d. PROJECT NUMBER
Rafael L. Bras, Jingfeng Wang	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Massachusetts Institute of Technology Office of Sponsored Programs Bldg. E19-750 Cambridge, MA 02139 -4307	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	50502-EV.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The objective of this proposal is to explore the idea of Maximum Entropy (MaxEnt) as a unifying principle to characterize multi-scaling hydrometeorological parameters such as rainfall and soil moisture, and develop a MaxEnt-based design of data collection networks for the sampling of multiscaling processes. The approach we follow is to relate the scaling distributions of hydrologic variables of interest to the macroscopic properties of the system and then use those distributions in a monitoring

15. SUBJECT TERMS

Maximum Entropy, Multi-scaling, Hydrometerology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Rafael Bras
UU	UU	UU	UU		19b. TELEPHONE NUMBER
					949-824-6002

Final Report of ARO W11NF-07-1-0126

Using the Maximum Entropy Principle as a Unifying Theory for Characterization and Sampling of Multi-scaling Processes in Hydrometeorology

Rafael L. Bras and Jingfeng Wang

25 February 2010

Activities and Findings

The project has focused on validating the theoretical results of characterizing multi-scaling processes in geomorphology using the Maximum Entropy Principle (MaxEnt) and developing a model of surface heat fluxes using the Principle of Maximum Entropy Production (MEP).

1. Validating MaxEnt Solution of Power-law Distribution

The power-law distribution of drainage area has been derived as the maximum entropy distribution under the constraint of given geometric mean. According to the Maximum Entropy theory, the power-law index, the parameter in the power-law distribution, is related to the geometric-mean of the drainage areas of a river basin and the lower limit of the drainage area that can be derived from digital elevation map (DEM). The agreement between the predicted values of the power-law index according to the MaxEnt with those obtained from the data as the curve-fitting parameters is an indicator that the MaxEnt can be used a general organizing principle in understanding and characterizing the multi-scaling processes in sampling network design.

It has been known that the drainage area for a river network, A, follows a power-law distribution,

$$P(A>a) \propto a^{-\delta} \tag{1}$$

where δ is the power-law index that can be determined from DEM data. The MaxEnt theory predicts that δ is related to the aggregated properties of the river network through,

$$\delta^{-1} = \ln\left(\frac{A_g}{A_1}\right) \tag{2}$$

where A_g is the geometric-mean and A_1 the lower-limit of the drainage areas of a river network. Figure 1 shows the DEM of Grand Canyon from USGS data archive. The river network is generated from the
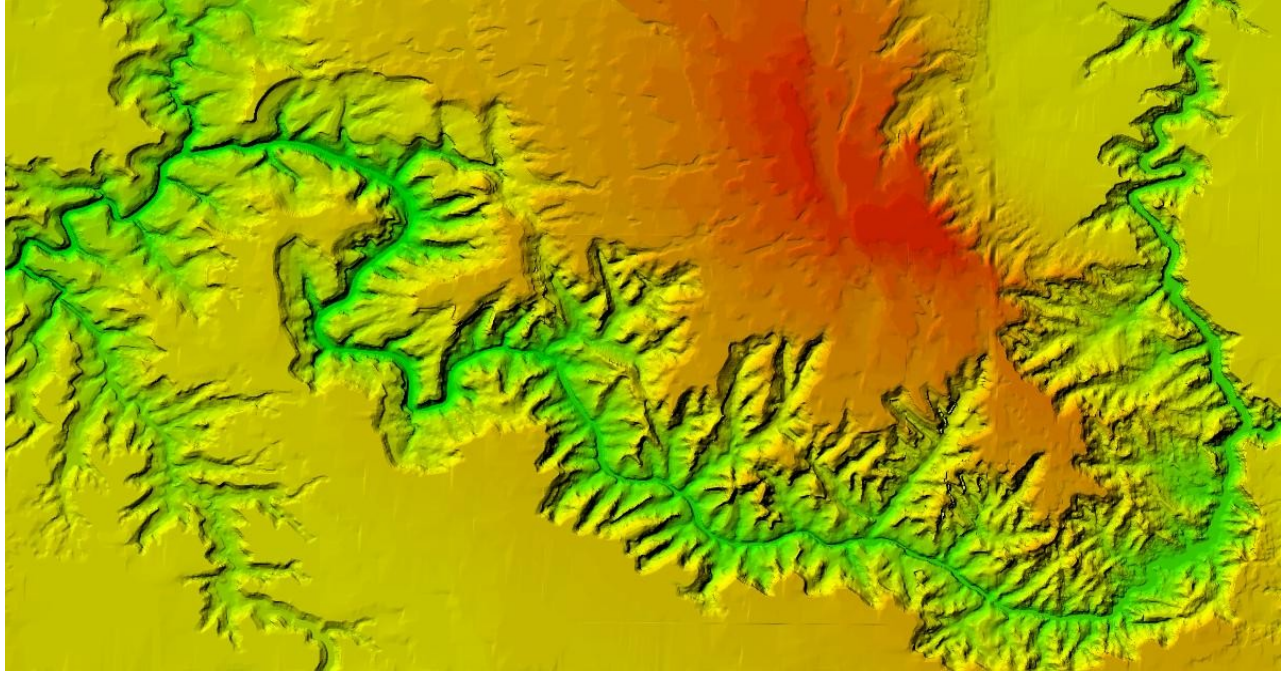
Figure 1. A DEM map for the Grand Canyon from USGS DEM datasets.
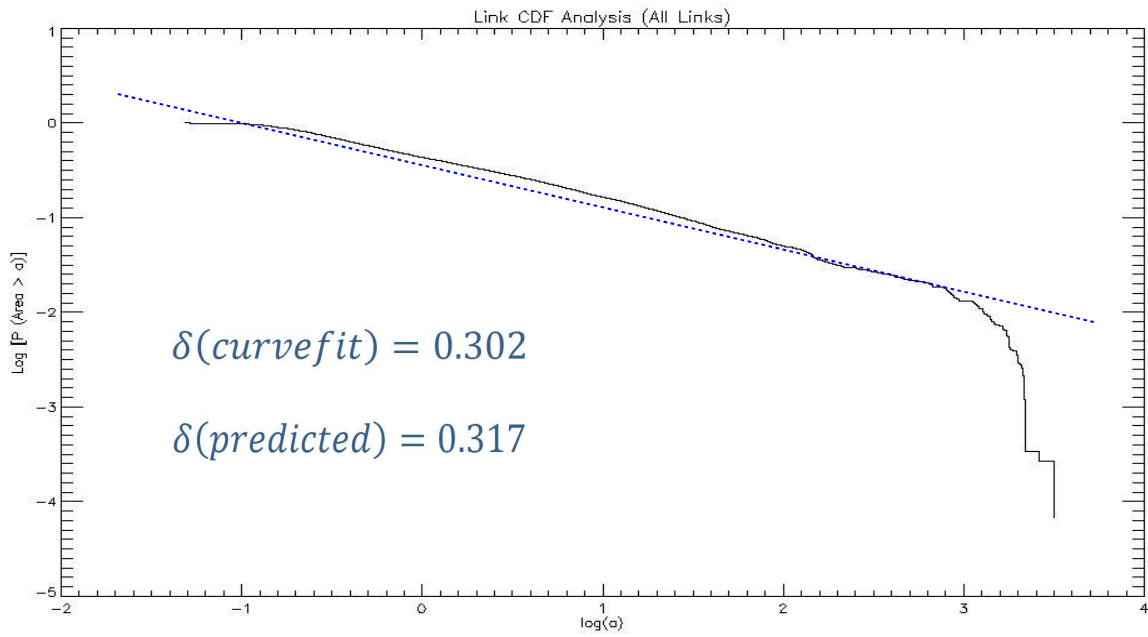
Figure 2. Computed cumulative distribution of the drainage areas using the DEM data shown in Figure 1. The predicted value of $\delta=0.317$ according to Eq (2), is in good agreement with the estimated $\delta=0.302$ obtained through curve-fitting the straight-line part of the empirical *CDF*.

Data Source	Predicted	Fitted	% Diff
Adjuntas	0.883	0.963	-8.36
Aguas Buenas	0.463	0.469	-1.37
Arecibo	0.514	0.512	0.46
Baldwyn	0.466	0.472	-1.27
Barceloneta	0.466	0.459	1.61
Barranquitas	0.487	0.495	-1.64
Bayamon	0.503	0.491	2.46
Bayaney	0.469	0.478	-1.85
Beaver Creek	0.390	0.403	-3.27
Caguas	0.470	0.482	-2.54
Camuy	0.469	0.486	-3.54
Carolina	0.474	0.528	-10.16
Cayey	0.488	0.492	-0.81
Central Aguirre	0.482	0.514	-6.22
Central La Plata	0.427	0.421	1.44

Table 1. Comparison of the parameters, δ, in the power-law distributions of drainage area (Eq(1)) using the MaxEnt theory with those estimated for the river networks in Puerto Rico derived from the USGS DEM data. "predicted" refers to the δ derived according to Eq (2), and "fitted" to the curve-fitting values of δ.

DEM data using a commercial software RiverTools. The computed cumulative distribution (*CDF*) of *A* is shown in Figure 2. The MaxEnt theory predicts $\delta=0.317$ according to Eq (2), while the empirical value of δ obtained from the curve-fitting of the straight line *CDF is 0.302,* which is in good agreement with the theoretical value of 0.317 according to the MaxEnt theory. Further tests of Eq (2) were done for several river networks in Puerto Rico using the USGS DEM datasets. The results are summarized in Table 1.

Of the fifteen river networks, twelve estimated parameters ("fitted" in Table 1) differ from the theoretical values ("predicted" in Table 1) by 1-3%. The largest difference is 10%. It is also found that the empirical values of δ is closer to the theoretical ones for larger river basins than for smaller river basins. This fact suggests that the some of the differences are related to the cut-off of the self-similar behavior of the systems at very small and very large scales (see Figure 2). For a wide range of scales at which the system manifests multi-scaling or self-similar behaviors, the MaxEnt principle correctly predicts the statistical properties of the system.

The validation study is the crucial first step in the study of *Type I* and *Type II* multi-scaling distributions for the proposed data-collection network design for sampling of multi-scaling processes in hydro-meteorology. The positive results have provided convincing evidence that the MaxEnt principle can be used as a unifying theory for characterization of self-similar processes. This finding affirms the basic ideas behind the project that the MaxEnt principle is able to *1*) assign probability distributions to the microscopic states of a system based on its macroscopic properties; *2*) retrieve useful information from

incomplete noisy data using entropy as a consistent quantitative measure of information. The finding

has wider scientific implications beyond our originally proposed study of sampling network design. For example, the analysis we did for the Grand Canyon and Puerto Rico can be done for the Martian surface using *M*ars *O*rbiter *L*aser *A*ltimeter (*MOLA*) data as the topographical features of Martian surface visually have some similarities with those of the earth. By computing the power-law index, if the drainage area for a Martian "river network" indeed follows similar distributions as shown here, we may be able to answer the challenging question whether the "river networks" on Mars were formed by flowing water like on the earth. As the most recent discovery of ice on Mars by the Phoenix Mars Mission, the MaxEnt tool developed here may have greater potential than we initially anticipated.

2. Specification of Non-informative Prior and the Information Entropy

The fundamental idea of the project is built on the expression of information entropy, *S*, that takes the form,

$$S = -\int f(x) \ln f(x)\, dx, \qquad (3)$$

where $f(x)$ is the probability density function of a continuous variable x. This is a special case of the general expression of information entropy under the condition of uniform probability measure $m(x)=1$,

$$S = -\int f(x) \ln \frac{f(x)}{m(x)}\, dx, \qquad (4)$$

where m(x) is not constant in general. *S* in Eq (4) is invariant with a change of variable, a desired property of information entropy. $m(x)$ is often referred to as ignorance prior or non-informative prior. It must be specified in order to obtain a solution of probability distribution using the framework of MaxEnt. An unanswered question is what functions can be used as ignorance priors? Use of the uniform ignorance prior, i.e. $m(x)=1$, has been criticized for: *1*) lacking invariance with change of variable, and *2*) improper (non-integrable) over an infinite domain. Since the formulation of the project assumes uniform ignorance prior, a justification is needed for the purpose of this project in particular, and for answering the question of ignorance priors in general.

Investigation of the problem has led to the following conclusions *1*) Ignorance prior is identical to the *l*imiting *d*ensity of *d*iscrete *p*oints (LDDP) function, and is not unique due to unlimited possibilities of re-parameterization (or change of variable). A LDDP function does not have to be proper; 2) The Principle of Indifference is a logical and sufficient rule for selecting ignorance priors applicable to discrete as well as continuous parameters. The Principle deprives any arbitrariness and ambiguity in selecting ignorance prior distribution; 3) A probability distribution defined on continuous parameter(s) consists of two terms: a LDDP function enumerating the hypotheses that is variant with re-parameterization and a probability function quantifying the likelihood of each hypothesis that is invariant with re-parameterization.

These results justify the use of uniform ignorance prior in the expression of information entropy. They have important theoretical implications. The concept of the LDDP facilitates the resolutions of some puzzles in the probability theory such as Borel-Kolmogorov paradox and Bertrand's problem. These results, if accepted, lay a sound foundation beneath the Bayesian analysis of a physical system with

incomplete information using the framework of MaxEnt.

3. Field Experiment

A field experiment at the Lucky Hill site (31° 44.564'N, 110° 3.251'W) in the Walnut Gulch watershed was conducted 2-17 June 2008 using the equipment acquired under the support of another ARO project W91NF-06-1-0224 to measure a full suite of surface energy balance and soil variables including downward and upward solar and long wave radiative fluxes. A eddy-covariance system was deployed to measure turbulent and conductive heat fluxes at the surface together with skin temperature and soil temperature at a certain depth. The experiment was during the pre-monsoon season characterized by strong insolation and occasional clouds. The experimental site is located at a mostly bare spot over a flat area. The soil does not contain removable moisture from the surface down to at least tens of centimeters. The sampling period followed an extensive period without rain. Skin temperature was measured using an infrared thermometer. All variables were sampled every 10 seconds. The collected data (Figures 3-4) will be used to understand the fundamental physics behind the partition of net radiation into turbulent and conductive heat fluxes from the perspective of MaxEnt and the maximum entropy production theory. The analysis of the dataset is underway.
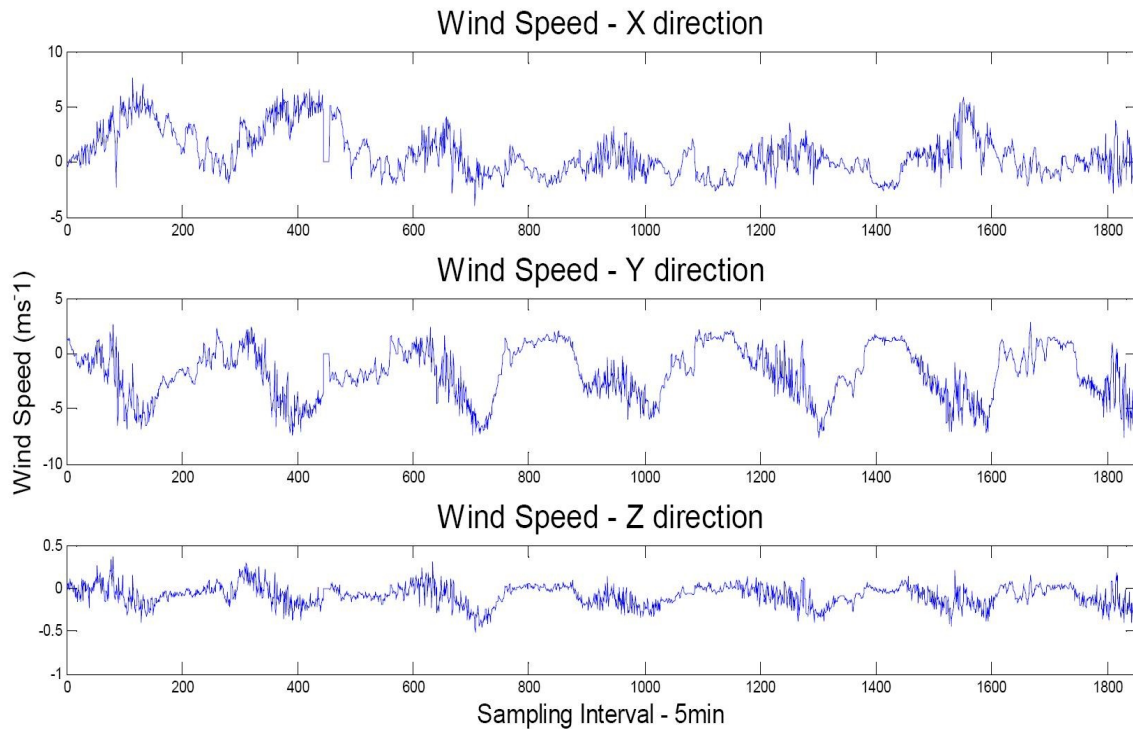
Figure 3. Wind velocities (sampled at 10 Hz) measured by the eddy-covariance device at the Walnut Gulch watershed during 9-15 June 2008.
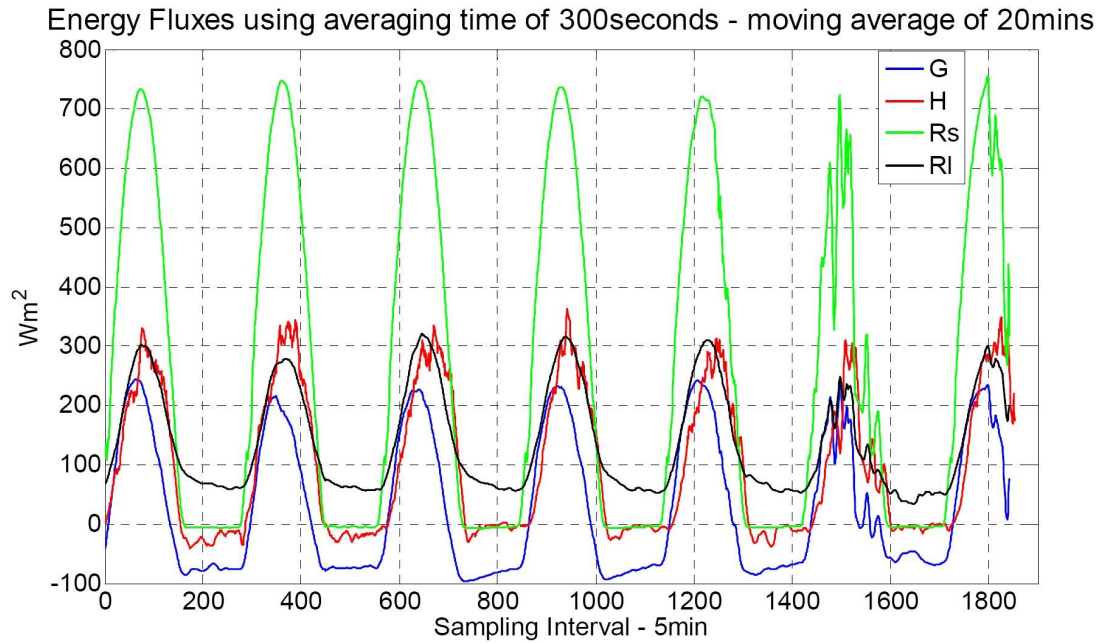
Figure 4. Surface energy fluxes (sensible, ground, up- and down-ward solar and long-wave) measured by the eddy-covariance device and radiometers at the Walnut Gulch watershed during 9-15 June 2008.

4. Development of a MEP Model of Surface Heat Fluxes

As a proof-of-concept study, a model of heat fluxes with a stationary hypothesis of the energy balance over a dry land surface are proposed based on the MEP theory. The first step is to formalute an entropy production function or dissipation function as a nonlinear function of the heat fluxes including "thermal inertia" parameters to represent the corresponding transport mechanism. The "thermal inertia" for the turbulent heat transfer in the atmospheric boundary layer is parameterized using a new solution of the Monin-Obukhov similarity model. A solution of the heat fluxes can be obtained by finding the extreme of the dissipation function under the constraint of conservation of energy for a given energy input (i.e. net radiation) at the surface. The MEP solution of the surface heat fluxes is tested using observations from fields experiments. A sample result is displayed in Figure 5 where the heat fluxes predicted by the MEP theory agree closely with the observations.

There are several important findings and conclusions from this study. The stationary hypothesis adds another example of natural phenomena understood from the perspective of an optimality argument. The MEP method for modeling and estimating surface heat fluxes complements the existing methods as it only needs net radiation input without requiring other meteorological variables such as temperature and wind speed. The MEP model, not replace the existing ones, offers an alternative method that would

enhance the models of land surface energy balance. The concept of entropy production is not limited to the thermodynamic entropy production. Depending on the quantities of interest, the "entropy (production)" defined in the MaxEnt and MEP formalisms is not always related to the thermodynamic entropy. This study presents an example where the "entropy production" is different from both thermodynamic entropy production and (thermal) energy dissipation that are often used in the models of non-equilibrium systems. The MEP method is expected to be applicable to the energy balance over a wet or a vegetated land surface.
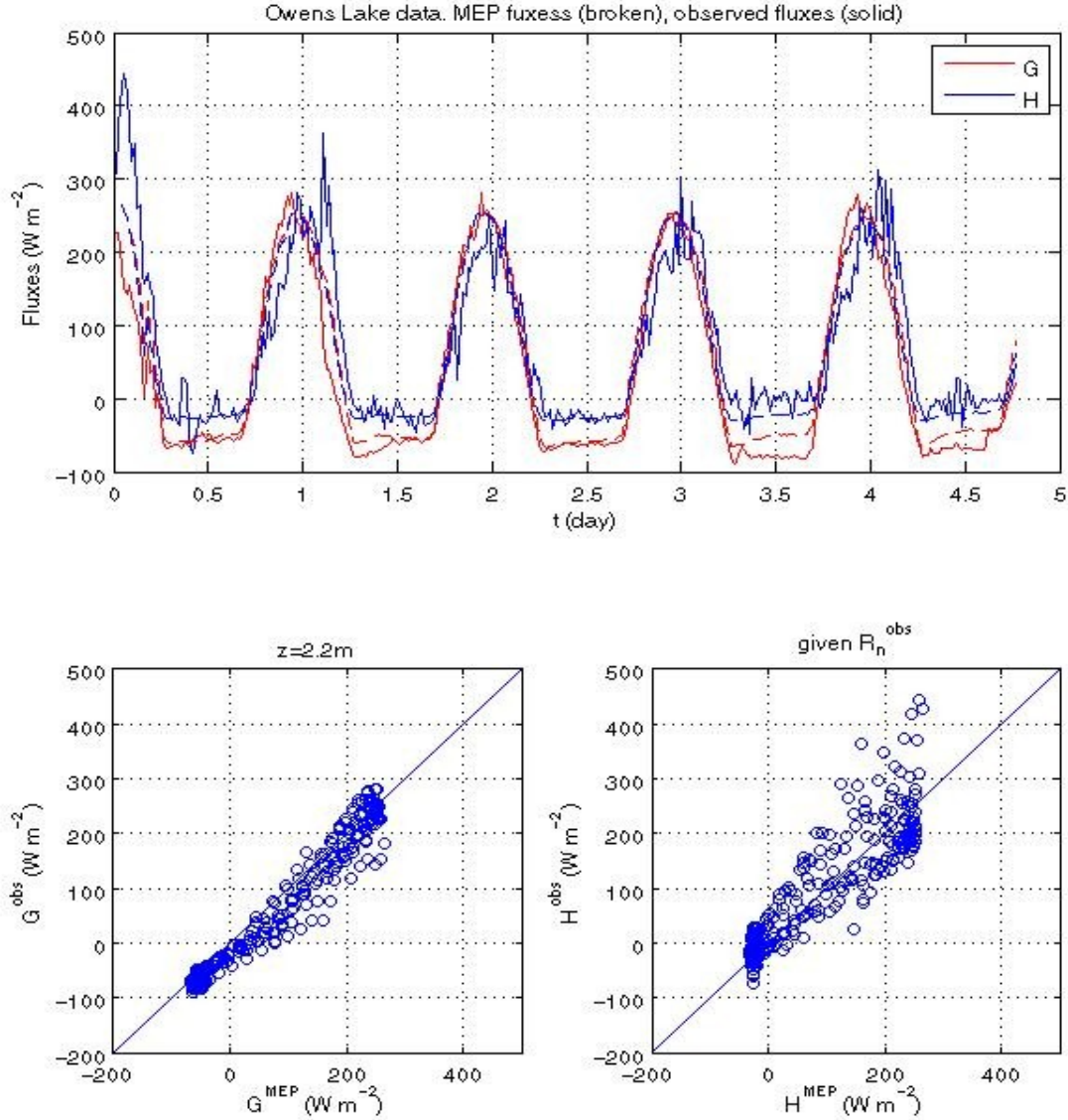
Figure 5. (a) Time-series of the MEP solution of G and H (broken lines) vs the observed fluxes (solid lines) with the observed Rn as the input to Eq (15); MEP solution of G (b) and H (c) vs observations.

20-minute average data collected at Owens lake, California, 20 June to 2 July 1993.

5. Turbulence of Atmospheric Boundary Layer

A new solution of the Monin-Obukhove similarity equations for the turbulence in the atmospheric boundary layer has been derived and tested, which plays a critical role in the development of MEP model of surface heat fluxes reported above. The major findings include the extremum solution of the Monin-Obukhov similarity equations: 1) is the only mathematically consistent and physically realistic solution of the M-O similarity equations; 2) has overcome some technical difficulties (e.g. nonuniqueness and non-convergence) in applying the MOST in modeling of turbulent transport in the ASL; 3) opens a possibility of simplifying the MOST formalism by replacing the two empirical stability functions by some empirical constants; 4) unifies the asymptotic solutions of the ASL derived from various arguments; 5) has a solid foundation built on the modern non-equilibrium thermodynamics; and 6) plays a crucial role in successful modeling of the surface heat fluxes based on the emerging theory of maximum entropy production. Our analysis also raises a doubt on the classical interpretation of the Obukhov length arguably due to a mathematical artifact. Resolution of the issue may offer new opportunities in improving the atmospheric turbulence models.

Journal Publications

1. Wang, J., and R. L. Bras, A model of surface heat fluxes based on the theory of maximum entropy production, *Water Resour. Res.*, 45, W11422, 2009.
2. Wang, J., and R. L. Bras, An extremum solution of Monin-Obokhov similarity equations, *J. Atmos. Sci.*, 67(2), 485-499, 2010.
3. Wang, J., R. L. Bras, G. Sivandran, and R. G. Knox, A Simple Method for the Estimation of Thermal Inertia, *Geophys. Res. Lett.*, 2010, in press.
4. Nieves, V., J. Wang, and R. L. Bras, The Maximum entropy principle as a unifying theory for characterizing self-similar processes, *Phys. Rev. Lett.*, 2010 (to be submitted).

Conference Presentations

1. Wang, J, and R. L. Bras, On Ignorance Prior Distributions and the Principle of Indifference, *Workshop on maximum Entropy Production in the Earth System*, Max-Planck Institute for Biogeochemistry, 07-09- May 2008.
2. Wang, J. E. Wood, and R. L. Bras, Maximum Entropy Solution of Power-law Distribution of River Networks, *AGU Fall Meeting*, San Francisco, 2008.
3. Wang, J. and R. L. Bras, Application of the Principle of Maximum Entropy Production (MEP) in Modelling Land Surface Energy Balance, *AGU Fall Meeting*, San Francisco, 2008.
4. Wang, J, and R. L. Bras, Application of the Theory of Maximum Entropy Production (MEP) in Hydrology – A Model of Surface Heat Fluxes, *Workshop on maximum Entropy Production in the Earth System*, Max-Planck Institute for Biogeochemistry, 18-20 May 2009.
5. Wang, J., R. L. Bras, G. Sivandran, and R. Knox, A Simple Method for Estimation of Thermal Inertia Using Remote Sensing Observations, *AGU Fall Meeting*, San Francisco, 2009.